QH506.E5
mc

**STIC-ILL**

**From:** Turner, Sharon
**Sent:** Tuesday, March 05, 2002 2:16 PM
**To:** STIC-ILL
**Subject:** 09292862

Please provide

Pierrou et al., EMBO J., 13:5002-12, 1994

Sharon L. Turner, Ph.D.
USPTO
CM1-10B09
Mailroom 10C01
Biotechnology GAU 1647
(703) 308-0056

The EMBO Journal vol.13 no.20 pp.5002–5012, 1994

# Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending

**Stefan Pierrou[1], Marika Hellqvist[1], Lena Samuelsson[1], Sven Enerbäck and Peter Carlsson[2]**

Department of Molecular Biology, Göteborg University, Medicinaregatan 9C, S-413 90 Göteborg, Sweden
[1]The first three authors contributed equally to this work
[2]Corresponding author

Communicated by K.Nordström

**The forkhead domain is a monomeric DNA binding motif that defines a rapidly growing family of eukaryotic transcriptional regulators. Genetic and biochemical data suggest a central role in embryonic development for genes encoding forkhead proteins. We have used PCR and low stringency hybridization to isolate clones from human cDNA and genomic libraries that represent seven novel forkhead genes, *freac-1* to *freac-7*. The spatial patterns of expression for the seven *freac* genes range from specific for a single tissue to nearly ubiquitous. The DNA binding specificities of four of the FREAC proteins were determined by selection of binding sites from random sequence oligonucleotides. The binding sites for all four FREAC proteins share a core sequence, RTAAAYA, but differ in the positions flanking the core. Domain swaps between two FREAC proteins identified two subregions within the forkhead domain as responsible for creating differences in DNA binding specificity. Applying a circular permutation assay, we show that binding of FREAC proteins to their cognate sites results in bending of the DNA at an angle of 80–90°.**

*Key words:* DNA bending/DNA binding/forkhead/transcription factor

## Introduction

The key event in gene regulation, control of initiation of transcription, depends on the coordinated activity of sequence-specific DNA binding proteins. Combinatorial effects generate independent regulation and cell type-specific expression for genes far more numerous than the transcriptional regulators themselves. Several mechanisms contribute to the complexity of transcriptional regulation by *trans*-acting factors. The formation of heterodimers between two DNA binding proteins can alter their ability to activate transcription, their affinity for DNA or sequence specificity and the stability of the dimer itself (Lamb and McKnight, 1991). Overlapping, yet distinct, binding site preferences among related transcription factors allow two promoters to utilize the same set of factors but with different relative affinities. Synergy or antagonism between transcription factors can act at the level of DNA binding (Grueneberg *et al.*, 1992) or transcriptional activation (Lin *et al.*, 1990; Herschlag and Johnson, 1993), and can make the activity of a certain DNA binding protein highly context-dependent (Carlsson *et al.*, 1993; Giese and Grosschedl, 1993). Alterations in DNA topology also generate context dependence; some regulatory proteins introduce sharp bends in the DNA, the effect of which depends on the position and orientation of the binding site (Natesan and Gilman, 1993; Grosschedl *et al.*, 1994).

The modular structure of eukaryotic transcription factors, where distinct functions such as DNA binding and transcriptional activation are often contained within non-overlapping protein domains, suggests that any class of DNA binding motif would be capable of mediating any kind of biological signalling. Nevertheless, regulatory proteins with the same basic design in their DNA binding domains also tend to be related in function, as seen in the steroid receptor superfamily or the homeobox proteins. This may reflect that evolution prefers to move in small steps and in creating a new entity will first look to the closest relative, but also that the structure of the DNA binding domain is not irrelevant for overall function.

The forkhead domain is a 100 amino acid motif that defines a rapidly growing family of DNA binding proteins. First identified as a region of homology between the product of the homeotic *Drosophila* gene *forkhead* and hepatocyte nuclear factor 3 (HNF3) from rat (Weigel and Jäckle, 1990; Lai *et al.*, 1991), the forkhead motif has since been found in genes from a number of metazoans and in *Saccharomyces*. Some of these genes have been isolated based on their homology to *forkhead* or *HNF3* and little is known about their function. This group includes five *forkhead*-related genes from *Drosophila* (*FD1–FD5*) (Häcker *et al.*, 1992), nine from rat (*HFH1–HFH7* and *HFH-B2* and *HFH-B3*; Clevidence *et al.*, 1993), six from mouse (*fkh1–fkh6*) (Kaestner *et al.*, 1993) and one from *Saccharomyces* (*HCM1*; Bork *et al.*, 1992), which was also independently isolated as suppressor of a calmodulin mutation (Zhu *et al.*, 1993).

Other members of the forkhead family have been identified as genes involved in pattern formation during embryogenesis. Members of this group include *forkhead* (Weigel *et al.*, 1989) and *sloppy paired* (*slp1* and *slp2*; Grossniklaus *et al.*, 1992; Häcker *et al.*, 1992) from *Drosophila*, *lin-31* from *Caenorhabditis elegans* (Miller *et al.*, 1993) and *Axial* from zebrafish (Strähle *et al.*, 1993). Indirect evidence suggests a similar function for a number of other forkhead genes. The embryonic expression pattern in mice implies that *HNF3α*, *HNF3β* and two related genes, *mf1* and *mf2*, are involved in the formation of the body axis and the establishment of the germ layers during gastrulation (Sasaki and Hogan, 1993). Ectopic expression of *HNF3β* in transgenic mouse embryos identified this gene as a regulator of floor plate development

(Sasaki and Hogan, 1994). *HNF3β* is believed to induce the expression of *sonic hedgehog* in the notochord, floor plate and forelimb buds of the developing embryo, and *sonic hedgehog* appears to preserve the expression pattern through activation of *HNF3β* (Echelard *et al.*, 1993; Riddle *et al.*, 1993; Sasaki and Hogan, 1994). Based on the spatial and temporal distribution of their expression and results of mRNA injections, the *Xenopus* genes *xfkh-1* and *pintallavis* appear to have functions similar to *HNF3β* in embryonic development (Dirksen and Jamrich, 1992; Altaba *et al.*, 1993). Additional forkhead genes have been identified as encoding factors that, like HNF3, bind to regulatory elements in mammalian genes which are expressed in terminally differentiated cells. Human T-cell leukaemia virus enhancer factor (HTLF; Li *et al.*, 1992b) and interleukin binding factor (ILF; Li *et al.*, 1991, 1992a) belong to this group.

The forkhead domain of HNF3γ bound to DNA has been crystallized and the 3-D structure determined (Clark *et al.*, 1993). The forkhead domain turns out to be a variant of the helix–turn–helix motif; it binds DNA as a monomer and contains two loops on the C-terminal side of the helix–turn–helix. Direct base contacts are made by the recognition α-helix that intrudes into the major groove of DNA and additional backbone contacts are provided by the loops. This variant of the helix–turn–helix with loops or wings that project over the DNA, has been given the name 'the winged helix' (Brennan, 1993; Clark *et al.*, 1993; Lai *et al.*, 1993).

We have shown previously that adipocytes contain a DNA binding activity with a specificity similar to that of HNF3 but not recognized by antisera against HNF3α or HNF3β (Enerbäck *et al.*, 1992). Regions in the promoter of the lipoprotein lipase (*lpl*) gene, which mediate the activation of *lpl* characteristic for differentiating adipocytes, contain binding sites for this activity. This observation indicated to us the existence of additional forkhead genes and their possible role in differentiation, and prompted us to try to isolate clones for such genes.

In this paper we describe the cloning of seven new members of the forkhead gene family from human cDNA and genomic libraries. To indicate the identity of their DNA binding domains and the fact that the ones tested in cotransfection experiments activate transcription (M.Hellqvist, L.Samuelsson, S.Pierrou, S.Enerbäck and P.Carlsson, manuscript in preparation), we named the proteins forkhead related activators (FREAC). The expression pattern for each *freac* gene was found to be unique and range from restricted to a single tissue to nearly ubiquitous. The DNA binding specificities of four of the FREAC proteins were determined through selection of high-affinity binding sites from random sequence oligonucleotides. All four proteins share a requirement for the core sequence RTAAAYA, but differ in their preferences 5′ and 3′ of the core, as well as with respect to the preferred nucleotides at the R and Y positions in the core. The results of swaps between two FREAC proteins point to a region at the N-terminal border of the recognition helix, helix 3, and another region at the C-terminal part of the forkhead domain, wing 2, as subdomains responsible for differences in DNA binding specificity. Using a circular permutation assay, we show that binding of the forkhead domain of FREAC proteins to their cognate DNA site results in bending of the DNA at an angle of 80–90°.

## Results

### *Cloning of human forkhead genes*

With the objective of identifying new members of the forkhead gene family, we adopted a PCR-based strategy with primers designed from regions conserved between rat *HNF3* and *Drosophila* forkhead, and a template consisting of cDNA made from human mRNA. PCR products of the expected size were cloned and sequenced. In addition to the human *HNF3* homologues, a clone was obtained that did not correspond to any known gene and with the potential to code for a forkhead protein. This clone was used to screen two human cDNA libraries and one human genomic library. From a fetal human cDNA library, *freac-1*, *freac-2* and *freac-3* were isolated. *freac-4*, which is identical to the clone generated by PCR, and *freac-5* were pulled out of a cDNA library made from the human monocyte cell line THP-1. Among the genomic clones isolated, two represented novel genes: *freac-7* contains the entire forkhead domain within a single exon, whereas *freac-6* is interrupted by an intron. A comparison of the predicted amino acid sequences of the seven FREAC proteins (Figure 1) reveals that FREAC-1 and FREAC-2 are nearly identical within the forkhead domain, although they show no similarity in primary structure in other parts of the proteins. The same relationship exists between FREAC-4 and FREAC-5.

### *Tissue distribution of expression*

Probes derived from the seven *freac* genes were used to probe Northern blots with RNA from multiple human tissues of both adult and fetal origin (Figure 2). The high degree of sequence conservation within the forkhead motif necessitated the use of probes derived from unique regions corresponding to other parts of the protein or untranslated sequences to avoid cross-hybridization. Almost identical patterns of expression are shown by *freac-1* and *freac-2*, with comparatively high levels of RNA in placenta and lung (adult and fetal). Following longer exposures, very low levels of expression of *freac-2* are seen in prostate, small intestine, colon and fetal brain. The expression of *freac-3* is nearly ubiquitous in adult tissues, with the highest levels detected in skeletal muscle, kidney, liver and heart. *freac-3* expression is absent from spleen, fetal brain and fetal lung, and very low levels are detected in testis, small intestine and leukocytes. Apparently, *freac-3* transcripts are differentially processed and sizes smaller than the dominating 3.9 kb mRNA are seen in fetal colon, leukocytes and fetal kidney. Among the tissues examined, expression of *freac-4* is detected exclusively in testis and kidney (adult and fetal), and the level of expression is several-fold higher in fetal kidney compared with adult. Since *freac-4* was isolated from a cDNA library made from THP-1 cells, we confirmed the expression of *freac-4* in these cells (data not shown), a result which implies that *freac-4* is also expressed in cells of the monocyte lineage. Muscle tissues heart and skeletal muscle are the principal sites of *freac-5* expression, but low levels of *freac-5* mRNA are seen in most tissues (placenta, thymus, fetal brain and fetal lung being the only exceptions). Of

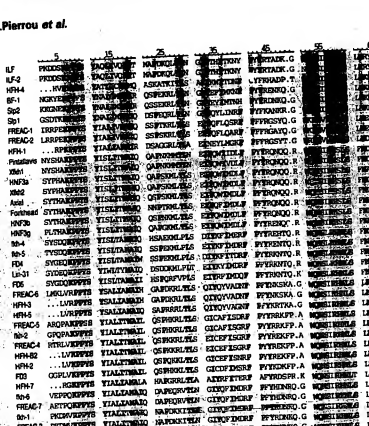

**Fig. 1.** The forkhead domains of FREAC proteins aligned to other known members of the forkhead family. Alignment and sorting according to relationship were performed using the Pileup program of the UWGCG software package (Genetics Computer Group, 1991). Shading indicates identity in >80% of the sequences. GenBank accession numbers for FREAC1–7 are U13219–U13225, respectively.

the *freac* genes, *freac-6*, is the most tissue-specific, with expression being detected only in kidney. The relative levels of mRNA between adult and fetal kidney are reversed compared with *freac-4*; in the case of *freac-6* the highest level is produced in adult tissue. No expression could be detected in any of the tissues investigated with a probe derived from *freac-7*. Hybridization with a β-actin probe verified the presence of intact mRNA in each lane of the blot.

### *Selection of binding sites*

The variation in primary structure within the forkhead domains of the FREAC proteins suggested that they may have different DNA binding specificities. To address this, we expressed the forkhead domains of four of the FREAC proteins in *Escherichia coli* and selected high-affinity binding sites with each one of them from a pool of random-sequence oligonucleotides. Since the pairs FREAC-1/FREAC-2 and FREAC-4/FREAC-5 are close to identical within their forkhead domains, we chose to determine the binding specificity of one representative from each pair (FREAC-2 and FREAC-4), in addition to FREAC-3 and FREAC-7. Recombinant FREAC proteins were expressed as fusions with glutathione S-transferase (GST), and the ability of GST to bind avidly to glutathione–Sepharose was used as a way of immobilizing the protein–DNA complexes. The oligonucleotides used for selection carried constant flanking sequences for PCR amplification and

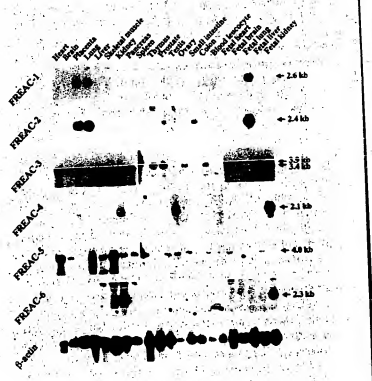

**Fig. 2.** Northern blots with human polyadenylated RNA analysed with probes specific for *freac-1* to *-6* and β-actin. No expression could be detected in the tissues investigated with a probe derived from *freac-7*.

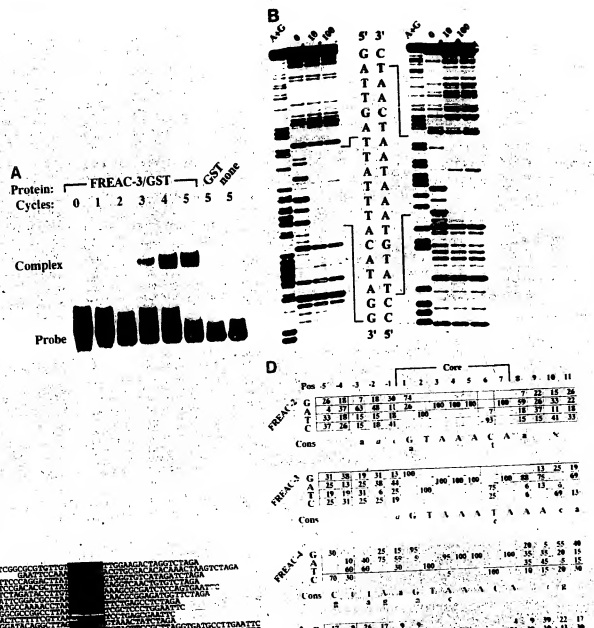


Fig. 3. Selection of binding sites from random-sequence oligonucleotides with FREAC–GST fusion proteins. (**A**) Enrichment of high-affinity binding sites following sequential selection and amplification: gelshift analysis of the PCR products obtained after each round of amplification (cycles 1–5) and the randomized oligonucleotide used as starting material for the selection (cycle 0). (**B**) DNase I footprinting on a sequence selected with FREAC-3–GST. Numbers above each lane indicate the amount of FREAC-3–GST protein in ng. (**C**) Sequences selected with FREAC-3–GST. The 32 nucleotides corresponding to the randomized segment are shown together with the first restriction site in each of the flanking sequences (*Eco*RI and *Xba*I). The core sequence of the FREAC-4 binding site is shaded. (**D**) Summary of selected binding sites for four FREAC proteins. Numbers represent the frequency percentage for each nucleotide at each position. The number of sequences on which this summary is based is: 27 for FREAC-2, 16 for FREAC-3, 20 for FREAC-4 and 23 for FREAC-7.

cloning, while the central 32 nucleotides were randomized. The enrichment of FREAC binding sites was followed by a gelshift assay, and after five or six rounds of selection and amplification a mixture of high-affinity binding sites was obtained (Figure 3A). For each FREAC protein, ~30 different sequences were compared and a consensus sequence calculated. The outcome is exemplified in Figure 3C by sequences selected with FREAC-4–GST. That the correct sequence had been identified was verified by DNase I footprinting, as shown for FREAC-3–GST in Figure 3B. Correlation between agreement with the consensus sequence and high affinity was confirmed with gelshift and quantitative filter binding assays (data not shown). To rule out that the GST moiety of the fusion protein has any affinity for the selected sequences, GST protein without anything fused to it was also tested in gelshifts with selected sequences; in no case was any interaction between GST and DNA observed. In addition to sequences that contain a single binding site within the 32 randomized nucleotides, we also found a second class

of selected sequences. These sequences contain sites that deviate more from the consensus and instead have two, or sometimes three, overlapping, sites. This suggests that our selection procedure was sufficiently stringent to retain only those oligonucleotides that bound avidly to the FREAC–GST fusion protein immobilized on Sepharose, but that the length of the randomized sequence allowed more than one site to contribute to binding. Hence, a fraction of the selected sequences produced the required binding strength through the combined action of two or three sites. To ensure that weak sites did not contribute, only monovalent sequences were included in the calculation of the final consensus summarized in Figure 3D.

### DNA binding specificity

A comparison of consensus sequences, generated through selection with the four different FREAC proteins, reveals that they share a requirement for the core sequence RTAAAYA. The positions within the binding site will be referred to relative to the first position of the core, i.e. the R position (Figure 3D). DNase I footprinting (Figure 3B) shows that the binding site is centred around the core; the DNA is protected from five nucleotides 5′ (position −5) to six nucleotides 3′ (position +13) of the core. Enhanced cleavage, indicating DNase I hypersensitivity induced by binding of the FREAC protein, is seen on the upper strand at the Y position (+6) and on the opposite strand at position +3. G is the preferred nucleotide at the R position (+1) for all the proteins except FREAC-7, which preferentially selected sites with an A in this position. In the Y position (+6), all the proteins except FREAC-3 preferred a C over a T. The consensus site for binding of HNF3 contains the RTAAAYA core with a single nucleotide difference (position +3): CTAAGTCAATA (Costa *et al.*, 1989). Examination of the structure of HNF3γ bound to DNA (Clark *et al.*, 1993) shows that the core sequence consists of the nucleotides where the recognition helix, helix 3, makes major groove contacts with DNA. Judged from the selected binding sites, a close agreement with the consensus within the core sequence positions is mandatory for high-affinity binding of FREAC proteins. No position within the core deviates from the RTAAAYA in >9% of the selected sequences for any of the FREAC proteins. In particular, the A doublet at positions +4/+5 appears to be critical since no exception from AA was observed. Outside the core, 3′ as well as 5′, the FREAC proteins exhibit different preferences, and variations in these flanking sequences appear to be better tolerated than within the core. To investigate the importance of flanking sequences on specificity, we synthesized oligonucleotides with different combinations of 5′, 3′ and core sequences and tested their binding in a gelshift assay with FREAC proteins synthesized by *in vitro* translation. Figure 4B and C illustrates the influence on binding by FREAC-3 of differences in the nucleotides flanking the core. Probes A and B share the same 5′ flanking sequence and core, but differ in their 3′ flanking sequences; probe B conforms to the FREAC-3 consensus, AACA, while the corresponding positions in probe A have the sequence GCAT. As predicted from the nucleotide frequencies of the selected sites, FREAC-3 binds better to probe B than to probe A. Probes B and E are identical within the core and both have the optimal 3′ sequence for FREAC-3 binding: AACA. They differ, however, in the 5′ flanking sequence where probe B has the sequence CTTAA and probe E AGGCC. FREAC-3 binds probe E with much lower affinity than probe B. This result shows that although the nucleotide frequencies in the positions 5′ of the core imply that any nucleotide could occupy any position, certain nucleotide combinations will severely impede binding of FREAC-3. Probe F, which combines the 5′ sequence of probe E with the 3′ sequence of probe A, fails completely to bind FREAC-3 in spite of its consensus core sequence, which confirms the importance of nucleotides on either side of the core for high-affinity binding.

### Domain swaps

When we compared the relative affinities of FREAC-3 and FREAC-4 for probes A and B we found that FREAC-4 has a reversed preference compared with FREAC-3 and binds better to probe A (Figure 4D). To determine which subdomains within the forkhead motif mediate recognition of different parts of the binding site, we expressed chimeric proteins that consist of various combinations between FREAC-3 and FREAC-4 (Figure 4A and D). These proteins, referred to as SWAP-1 to SWAP-8, were translated *in vitro* and assayed for binding to four different probes. Probes A and B have been described above and differ in the four nucleotides immediately 3′ of the core (+8 to +11). Probes C and D differ only in the Y position of the core (position +6): C in probe C and T in probe D. As discussed above, FREAC-3 binds probe B with higher affinity than probe A, while the reverse is true for FREAC-4. Of the chimeric proteins, SWAP-1 to SWAP-4 have the same preference as FREAC-3, while SWAP-5 to SWAP-8 behave like FREAC-4 (Figure 4D). These results suggest that amino acids close to the C-terminus of the forkhead domain determine the specificity of each protein with regard to nucleotides 3′ of the core. In HNF3γ the corresponding region comprises the C-terminal half of wing 2 (Figures 4A and 6) which is dominated by a stretch of basic amino acids. Within this stretch, three residues differ between FREAC-3 and FREAC-4, two of which are conservative: R---KK in FREAC-3 versus K---RQ in FREAC-4 (Figures 1 and 4A). These residues define the C-terminal border of the forkhead homology and beyond this point the amino acid sequences of FREAC-3 and FREAC-4 diverge completely. The proteins used in binding experiments extend into the unique sequences on the C-terminal side of the forkhead domain by five and 16 residues respectively, and it is possible that amino acids in this region influence binding specificity. Probes C and D were bound equally well by FREAC-3, SWAP-1, -6, -7 and -8, while FREAC-4 and the remaining SWAP proteins bound better to probe C (Figure 4D). Preference for C at the Y position in the core therefore appears to be encoded by a region in the central part of the forkhead domain. The only differences in primary structure between FREAC-3 and FREAC-4 within the implied segment occur in a block of eight amino acids: F--DNKQG in FREAC-3 versus Y--EKFPA in FREAC-4 (Figures 1 and 4A). On the supposition that the basic structure is the same for all forkhead proteins, comparison with the 3-D structure of HNF3γ shows that the eight amino acids where the differences occur are located in

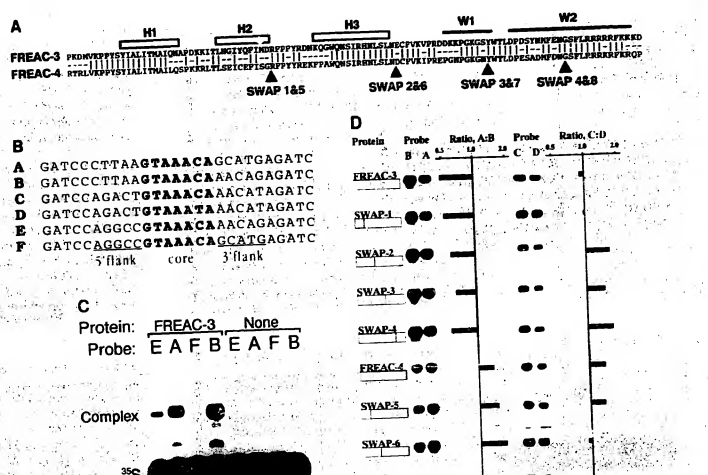


Fig. 4. Specificity of DNA binding by FREAC proteins. (**A**) Primary structures of the forkhead domains of FREAC-3 and -4. Breakpoints in the various SWAP proteins are shown as arrowheads under the amino acid sequences. Regions that correspond to the three α-helices (H1–H3) and the two wings (W1 and W2) in HNF3γ (Clark *et al.*, 1993) are indicated above the sequences. (**B**) Probes used in the gelshift assays shown in (C) and (D). (**C**) Gelshift assay with four of the probes in (B) and the forkhead domain of FREAC-3 *in vitro*-translated in a reticulocyte lysate. In the lanes marked 'Protein: None', a mock-translated reticulocyte lysate was used. '35S' indicates non-specific bands derived from [35S]methionine in the *in vitro*-translation reactions. (**D**) Gelshift assays with FREAC-3, FREAC-4 and chimeric proteins SWAP-1 to -8. The retarded bands, corresponding to protein–DNA complexes, are shown and the bar graphs indicate the ratio between the amount of complex formed with probe A versus probe B, or probe C versus probe D. For the exact design of the chimeric proteins see (A), and for probe sequences see (B).

the loop between helices 2 and 3 and in the first three amino acids of helix 3 (Figures 4A and 6).

### *DNA bending*

To investigate if binding of FREAC proteins affects DNA topology, we performed a circular permutation assay (Wu and Crothers, 1984). Oligonucleotides containing binding sites for FREAC-3 and FREAC-4 were cloned into a vector between two tandem copies of a 375 bp fragment. Digestion with restriction enzymes generated gelshift probes identical in size and sequence but with the FREAC binding site at different positions within the probe. Figure 5 shows the result of a gelshift with FREAC-3–GST and probes containing a FREAC-3 site at a variable distance from the end of the probe. The retarded complexes, representing FREAC-3 bound to DNA, migrate with a mobility that is inversely correlated to the distance between the binding site and the end of the probe, a relationship characteristic of proteins that bend their target DNA (Wu and Crothers, 1984). We repeated this assay on polyacrylamide gels with acrylamide concentrations of 6, 8 and 10% and calculated the ratio between the fastest and slowest migrating species for each gel concentration. These ratios were then used to estimate the extent of DNA distortion through linear interpolation between values obtained with A-tract DNA standards (Thompson and Landy, 1988). Independent of the gel concentration used, the angle of the DNA bend induced by binding of FREAC-3 was calculated to be between 80 and 90°. Consistent results were obtained with four different probes (A, B and E in Figure 4B, and G described in Materials and methods) and with FREAC-4 as well as FREAC-3. In agreement with the gelshift results presented above, probe F failed to bind FREAC-3 or FREAC-4 even when cloned into the circular permutation vector. In contrast to the other sequences tested, probe F appears to have a slight intrinsic curvature, as shown by differences in mobility of the free DNA (data not shown).

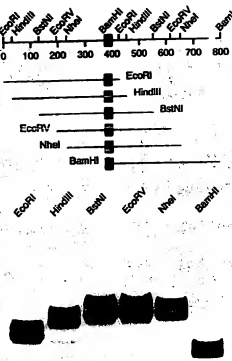


**Fig. 5.** Circular permutation assay. An oligonucleotide containing an optimal FREAC-3 site, probe G described in Materials and methods, was cloned between two tandem repeats of a 375 bp fragment and gelshift probes were generated through digestion with the indicated restriction enzymes. The FREAC-3 binding site is represented by the black rectangle. Gelshift assays were performed with FREAC-3–GST fusion protein and the mobilities of the fastest (*Bam*HI) and slowest (*Eco*RV) complexes were used to calculate the DNA bending angle. Results obtained with a 6% polyacrylamide gel are shown.

## Discussion

We have isolated clones for seven new members of the forkhead gene family from human DNA libraries. Sequence alignments of the seven *freac* genes, together with published forkhead genes, show that the FREAC proteins conform to the general pattern within the family of highly conserved amino acid motifs interleaved with more variable regions (Figure 1). In the cases of FREAC-1/FREAC-2 and FREAC-4/FREAC-5, the primary structures within the forkhead domains are almost identical between each pair, while amino acid sequences on either side of the forkhead homology show no or little resemblance. This may reflect that the proteins are designed to bind the same set of sequences, but otherwise have distinct functions. Alternatively, the proteins may be functionally redundant and the activities exerted by regions outside the DNA binding domain may have a greater tolerance with regard to amino acid substitutions, a condition that would explain the divergence in primary structure seen in these parts of the proteins. Such apparent flexibility in primary structure requirements is often seen in the mutational analysis of transcriptional activation domains; however, this explanation fails to account for the high degree of sequence conservation between species, generally found among transcription factors.

In the case of FREAC-1/FREAC-2, the homology between their DNA binding domains parallels a similarity in tissue distribution of expression. Both genes are expressed at fairly high levels in lung and placenta and it seems reasonable to assume that their target genes are the same. The forkhead motifs of *freac-1* and *freac-2* are not closely related to any other family member and these two genes appear to form their own subgroup. In contrast, the predicted amino acid sequence of FREAC-3 is identical within the forkhead domain to that of fkh-1 (Kaestner *et al.*, 1993) and frkhda (Sasaki and Hogan, GenBank accession number L10406). *fkh-1* and *frkhda* have both been cloned from mouse, are identical throughout, even at the nucleotide level, and therefore appear to be derived from the same gene. The sequence of *fkh-1* outside the forkhead domain has not been published; for *frkhda*, however, some additional sequence is presented and a comparison with *freac-3* shows that no similarity exists outside the forkhead homology. Furthermore, *fkh-1* is expressed in brain, heart, kidney and fat, while no expression is detected in skeletal muscle. *freac-3*, on the other hand, has its main site of expression in skeletal muscle, whereas expression in brain is hardly detectable. This suggests that *freac-3* represents a novel gene, while *fkh-1* and *frkhda* are derived from a gene whose human homologue remains to be cloned. A similar relationship exists between *freac-4* and *HFH-B2* (Clevidence *et al.*, 1993). Within the forkhead domain the predicted amino acid sequences are identical but the expression patterns are distinct. *freac-4* expression is specific for kidney and testis, while *HFH-B2* is reported to be expressed exclusively in brain. Evaluation of how similar these two genes are in regions other than the DNA binding domains awaits more sequence information on *HFH-B2*. *freac-4* and *HFH-B2* belong to a larger group of genes with closely related sequences within their forkhead motifs. This group includes *freac-5*, *HFH-6* (*fkh-2*), *HFH-2* and *FD3* (Häcker *et al.*, 1992; Clevidence *et al.*, 1993; Kaestner *et al.*, 1993).

*freac-6* appears to be the human homologue of *HFH-3* from rat (Clevidence *et al.*, 1993). Not only are the amino acid sequences within the forkhead domains identical, but expression of both genes is restricted to kidney. *freac-7* is most closely related to *fkh-6* from mouse (Kaestner *et al.*, 1993) and, based on the limited sequence information available, the homology appears to be confined to the forkhead motif. Examples of groups or pairs of genes encoding proteins with identical or very similar DNA binding domains thus appear to be common in the forkhead family, a phenomenon that was first illustrated by the isolation of the three *HNF3* isoforms α, β and γ (Lai *et al.*, 1991). Alternative splicing could generate distinct proteins with identical DNA binding domains from the same gene if exon borders coincided with the boundaries of the DNA binding domain. However, preliminary analysis of genomic clones for several *freac* genes (data not shown), as well as the gene structures of *HNF3*α, β and γ (Kaestner *et al.*, 1994), provide evidence against this hypothesis and we therefore conclude that strong selection pressures exist that control even minor variations within the forkhead domain.

We have determined the binding site specificities of four FREAC proteins. A core sequence, RTAAAYA, is common for binding sites selected with all four FREAC proteins, but each FREAC protein has its own signature with regard to the preferred nucleotides at the R and Y positions of the core. However, the flanking sequences appear to be more important in giving each protein its specificity, while at the same time being less well defined

than the core. A similar situation is seen in the family of homeobox proteins which share a requirement for the core sequence TAAT and where specificity is conveyed by nucleotides outside the core (Ekker *et al.*, 1991, 1992; Catron *et al.*, 1993).

Gelshifts with FREAC-3 and a set of probes (which all contained an intact core sequence) clearly demonstrated the importance of the flanking sequences. In spite of the fact that sites selected with FREAC-3 have all nucleotides represented at each of the five positions 5′ of the core, alterations in this part of the binding site had a dramatic impact on binding. HNF3γ interacts with the DNA 5′ of the core mainly through backbone contacts (Clark *et al.*, 1993). It seems likely that the general way in which HNF3γ binds DNA is relevant for the entire family of forkhead proteins. Interactions with the DNA backbone may be indirectly dependent on base sequence through effects on the topology and helicity of DNA, but less sensitive to single nucleotide substitutions than the direct base contacts made in the major groove of the core.

It is likely that some of the restraints on the flanking sequences of the binding sites emanate from the dramatic bending of DNA induced by binding of a FREAC protein. It should be emphasized though that the strong FREAC binding sites have no intrinsic curvature. In the circular permutation assay all probes tested migrate with the same mobility when present as free DNA.

The functional importance of DNA bending is best understood in prokaryotes. Bacterial regulatory proteins that bend DNA include transcriptional regulators such as CAP/CRP and architectural proteins exemplified by integration host factor (IHF; Wu and Crothers, 1984; Thompson and Landy, 1988; Moitoso de Vargas *et al.*, 1989; Zinkel and Crothers, 1991). IHF bends DNA by ~140° and facilitates interaction between proteins bound to distant sites by looping out the intervening DNA. Transcriptional activation at a distance by AraC (Schleif, 1992) and chromosomal integration of phage lambda (Moitoso de Vargas *et al.*, 1989) are examples of processes dependent on IHF-mediated DNA bending.

Among eukaryotic proteins the most severe distortion of DNA is generated by the HMG domain proteins, of which the best studied is lymphoid enhancer factor (LEF; Travis *et al.*, 1991; Waterman *et al.*, 1991). LEF has been proposed to act as an architectural protein in the assembly of enhancer nucleoprotein complexes (Giese *et al.*, 1992; Grosschedl *et al.*, 1994), but its mode of action is clearly distinct from that of IHF. DNA binding/bending is not sufficient for the activity of LEF and its context-dependent activation domain can be successfully grafted onto a heterologous non-bending DNA binding domain (Carlsson *et al.*, 1993; Giese and Grosschedl, 1993). In contrast, YY1, a zinc finger protein, appears to fulfil the criteria of an architectural protein. It acts as a repressor or an activator of transcription depending on the position and orientation of its binding site (Natesan and Gilman, 1993). The mechanism by which YY1 regulates transcription is through other proteins binding at flanking sites and it can be functionally replaced by an unrelated DNA bending protein.

Bending of DNA by members of the forkhead family has not been reported previously. We investigated how two of the FREAC proteins, FREAC-3 and FREAC-4, interact with four different binding sites, and found the angle of the DNA in complex with protein to be between 80 and 90° in each case. Based on the high degree of conservation within the forkhead domain (Figure 1), we predict that bending of the target site is an intrinsic characteristic of this class of DNA binding proteins.

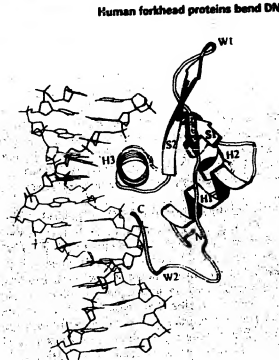


**Fig. 6.** The 'winged helix' structure of HNF3γ bound to DNA based on X-ray crystallography, as described by Clark *et al.* (1993). The three α-helices (H1 – H3), the two wings (W1 and W2), the three β-sheets (S1 – S3) and the N- and C-termini are indicated. The recognition helix (H3) is seen fitting into the major groove of the DNA in the region of the binding site which corresponds to the core sequence. The second wing (W2) contacts DNA in the minor groove on the 5′ side of the core sequence. Illustration adapted from Clark *et al.* (1993).

A 13mer oligonucleotide cocrystallized with HNF3γ has a curvature of only 13° (Figure 6; Clark *et al.*, 1993). This may reflect a difference in binding characteristics between on the one hand FREAC-3 and -4, and on the other HNF3γ. Alternatively, the discrepancy may reflect the different experimental methods applied. More work will be required to resolve the exact nature of complexes between forkhead proteins and DNA in solution, but two observations indicate that the X-ray structure could underestimate the degree of DNA distortion. First, the oligonucleotide used for crystallization does not extend beyond the 3′ border of the core. Secondly, the two last base pairs in the oligonucleotide cocrystallized with HNF3γ, which correspond to the two last positions of the core, differ from those in the HNF3 consensus and from the binding site in the transthyretin promoter on which the sequence was based (Costa *et al.*, 1989). Consequently, conceivable interactions or DNA distortions 3′ of the core were impossible to detect and contacts between DNA and protein in the 3′ end of the core may have been suboptimal. Inspection of the crystal structure shows that the first wing, W1, of HNF3γ projects beyond the 3′ end of the oligonucleotide and would be in a perfect position to make additional contacts with DNA had a 3′ flanking sequence been present. It is easy to envisage how the two wings, W1 and W2, could provide the interactions that

would bend the DNA, narrowing the major groove around the recognition helix, H3, in the process. Maybe the loops of the forkhead domain, rather than being wings as of 'a butterfly perched on a straight rod' (Brennan, 1993), are the arms of a brawny beast warping the DNA.

From gelshift assays performed with chimeric proteins between FREAC-3 and FREAC-4, we were able to pinpoint two subregions within the forkhead domain that influence binding site preferences. The relative affinities for T versus C at the Y position (+6) are determined by a short stretch of amino acids at the junction between the T' loop and helix 3. None of the DNA–protein contacts identified in HNF3γ will readily explain this connection, although the base pair that would be involved is one of the two discussed above that deviate from the HNF3 core consensus. However, the N-terminus of helix 3 and the T' loop are in close proximity to the base pair at position +6, which makes the conclusions from the domain swapping experiments seem logical.

Much more surprising is the outcome of the comparison of two probes with different sequences in the flanking region 3′ of the core. From the way HNF3γ interacts with its binding site, the first wing, W1, appears to be in the best position to interact with the 3′ flanking DNA, and we anticipated the primary structure of this subdomain to decide the preference of each chimera. Interestingly, the relative affinities of the chimeric proteins follow the origin of the second wing, W2, and non-conserved sequences C-terminal of the forkhead domain. W2 of HNF3γ contacts the DNA backbone in the minor groove on the 5′ side of the core. Hence, the way W2 influences DNA–protein interactions 3′ of the core is likely to be indirect, or alternatively residues outside the forkhead domain on the C-terminal side could contribute to DNA binding.

During the preparation of this article, DNA binding specificities of HNF3, HFH-1 and HFH-2 were published by Overdier *et al.* (1994). A stretch of 20 amino acids from the middle of helix 2 to five residues into helix 3 was found to contribute to differences in specificity between HFH-1 and HNF3β. This region encloses the eight amino acids around the N-terminus of helix 3 that we have identified as the determinant of the preference at nucleotide position +6 in the binding site.

In conclusion, subtle differences in the nucleotide sequence within and flanking the core generate diversity with regard to binding specificities of forkhead proteins. The amino acids that make direct base contacts are highly conserved throughout the forkhead family, an evolutionary preservation matched by the contacted nucleotides in the core of the binding sites. Indirect manifestations of the base sequence, such as DNA helicity and topology, appear to be important for the interactions that create diversity, interactions that are likely to include DNA backbone contacts. Amino acids around the N-terminal border of helix 3 and in wing 2 of the forkhead domain determine at least part of this specificity. A more thorough analysis of the structure of forkhead domains that represent distinct sequence specificities will be necessary to understand the way this large family of transcription factors manages to independently regulate its target genes. It will also be important to establish if bending of DNA is a general characteristic of forkhead proteins and, if so, how this ability influences their function as gene regulators.

## Materials and methods

### *Reagents and materials*

Enzymes, except for sequencing, were obtained from Boehringer Mannheim. Fluorescent DNA sequencing was performed with reagents from Pharmacia and oligonucleotides were synthesized on a Beckman Oligo1000. cDNA synthesis and cloning kit was purchased from Stratagene. QUICK-Clone cDNA, multiple-tissue Northern blots and pre-made libraries were acquired from Clontech. $^{32}$P-labelled nucleotides and [$^{35}$S]methionine were obtained from Amersham. Oligo(dT) Dynabeads were purchased from Dynal, rabbit reticulocyte lysate from Promega, glutathione–Sepharose and poly(dIC) from Pharmacia; Spin-X from Costar and MetaPhor agarose from FMC.

### *Isolation and sequencing of cDNA and genomic clones*

The primers 5′-GCTCATCACCATGGCCATCCAGCAG and 5′-CTTGAAGCCCTTTTGACGGCGCAAG were used to amplify forkhead motifs in a PCR with QUICK-Clone cDNA made from human adipocyte RNA as template. Conditions for the PCR were: 95°C, 1 min; 56°C, 2 min; 72°C, 3 min; 30 cycles. Products of the expected size were cloned and sequenced and a PCR product, whose sequence showed that it was derived from a previously unknown gene encoding a putative forkhead protein, was used to screen human cDNA and genomic libraries.

A cDNA library was constructed in the vector λZAPII, starting with 5 μg poly(A)$^+$ RNA prepared from the human monocyte cell line THP-1 (Tsuchiya *et al.*, 1980). RNA was prepared from THP-1 cells according to Chirgwin *et al.* (1979) and poly(A) selection was performed with oligo(dT) Dynabeads.

*freac-1*, *freac-2* and *freac-3* were isolated from a fetal human λgt11 cDNA library, *freac-4* and *freac-5* from the THP-1 cDNA library and *freac-6* and *freac-7* from a human genomic λ DASH library. All screenings were made with the PCR-derived probe, labelled with [α-$^{32}$P]dATP and post-hybridization washes were carried out at reduced stringency.

Nucleotide sequences were determined on a Pharmacia A.L.F. sequencer using T7 polymerase and either fluorescein-labelled primer or fluorescein–dATP.

### *Northern blots*

For each gene, a unique probe located outside the conserved region encoding the forkhead domain was used to probe Northern blots with poly(A)$^+$ RNA from multiple human tissues. Probes were labelled with [α-$^{32}$P]dATP, filters were hybridized at 47°C in 50% formamide and washed at full stringency (65°C, 0.1× SSC). Exposures ranged from overnight up to 1 week.

### *Bacterial expression*

DNA fragments encoding the forkhead domains of *freac-2*, *freac-3*, *freac-4* and *freac-7* were amplified with the following PCR primers: 5′-GGGAATTCCGGGGCTGCGGCGGCCCGAG, 5′-GGGGTCGACTTGAGCGCCTGGCACTTCCG (*freac-2*); 5′-GGGAATTCCTTACACGCCGCAGCCGCAG, 5′-AAAAGTCGACTCCTTGAGGTGCAGCCTGTC (*freac-3*); 5′-GGGGAATTCCGCCAAGAACCCGCTGGTG, 5′-GGGGGTCGACGGGTGGGAGCAGCGGCTGCC (*freac-4*); and 5′-AAGAATTCCCTCGGGCCGGGCCGAGACCCCGCAG, 5′-ATATGTCGACCTCCGGGGCCCCGGGGCCCGG (*freac-7*). PCR products were digested with *Eco*RI and *Sal*I, cloned between the corresponding sites in pGEX-KG (Guan and Dixon, 1991) and their authenticity verified by sequencing. Cultures of *E.coli* DH5α harbouring the respective pGEX-KG/FREAC plasmids were induced with 0.5 mM IPTG at an $OD_{550}$ of 0.3–0.5. The heating was turned off and the cultures were allowed to reach room temperature under vigorous shaking over a period of several hours. Bacteria were collected through centrifugation and gently resuspended in ice-cold TNT (10 mM Tris–Cl, pH 8.0; 1 mM EDTA; 100 mM NaCl; 1% Triton X-100). Approximately 1 mg of lysozyme was added; the suspension was immersed in ice water in an ultrasonic bath and sonicated for 5 min. The lysed bacteria were transferred to SW55 ultracentrifuge tubes and the solution cleared by centrifugation at 50 000 r.p.m. at 2°C for 15 min. GST/FREAC-7 was insoluble and had to be recovered from the pellet with 6 M guanidine–HCl, followed by renaturation through dialysis against TNT + 2 mM DTT and recentrifugation in SW55. Glycerol was added to the cleared lysate to 15%, DTT to 2 mM and aliquots were frozen in liquid nitrogen. GST/FREAC proteins were affinity purified on glutathione–Sepharose and eluted with 5 mM glutathione in 50 mM Tris–Cl, pH 8.0. Glycerol and DTT were added and aliquots frozen as described above.

### *Selection of binding sites*

A 72mer oligonucleotide with the 32 central bases completely degenerate [5′ - CGCTCGAGGGATCCGAATTC(N$_{32}$)TCTAGAA^GCTTGTCGA-CGC] was synthesized and purified with trityl by reverse-phase chromatography. 6 μg of this 72mer oligonucleotide were mixed with 5 μg of *Sal*I – *Xba*I primer (5′-GCGTCGACAAGCTTTCTAGA) in 10 mM Tris–Cl, pH 7.5, 10 mM $MgCl_2$ and 1 mM DTT, heated to 95°C and cooled slowly to 37°C. Cold dNTPs, 10 μCi [α-$^{32}$P]dATP and 5 U Klenow enzyme were added and the reaction was incubated at 37°C for 30 min. Successful conversion to double-stranded 72mer was verified by running an aliquot on polyacrylamide gel. The 72mer, theoretically 12 μg, was precipitated and resuspended in 100 μl 1× BB (20 mM HEPES, pH 7.9; 50 mM KCl; 2 mM $MgCl_2$; 0.5 mM EDTA; 10% glycerol; 0.1 mg/ml bovine serum albumin; 2 mM DTT; 0.5 mM phenylmethylsulfonyl fluoride) containing 10 μg poly(dIC). Total bacterial protein extract containing GST/FREAC protein was added and the binding reaction was incubated for 15 min at 25°C. The amount of bacterial extract added varied depending on the expression level of the GST/FREAC protein and was adjusted to correspond to ~10 ng of fusion protein. 50 μl of a 1:10 slurry of glutathione–Sepharose in 1× BB were added, the tube was flicked gently for 2 min and the Sepharose collected by a 1 min centrifugation at 5 kr.p.m. The supernatant was aspirated and the pellet washed four times with 1 ml 1× BB, changing tubes at the first and last washes. The washed Sepharose pellet was resuspended in 100 μl 1× PCR buffer containing 1 μM each of *Sal*I – *Xba*I primer and *Xba*I – *Eco*RI primer (5′-GCGTCGACAAGCTTTCTAGA) and 0.2 mM dNTPs. Amplification was performed as follows: 96°C, 1 min; 60°C, 30 s for 30 cycles followed by 1 cycle at 72°C for 30 s. An aliquot was checked on a 4% MetaPhor agarose gel; the rest of the PCR was precipitated, resuspended in 120 μl $H_2O$ and filtered through a 0.45 μm Spin-X. The binding reaction for the next round was identical to the first, except that 10 μl of the Spin-X filtrate was substituted for the 12 μg of random sequence 72mer. Five or six rounds of binding and amplification were made for each GST/FREAC protein. The number of cycles in the PCR were decreased successively as more of the target DNA was retained on the Sepharose, from 30 after the first round to typically 10 after the fifth. A small amount of the PCR product from each round was labelled with [γ-$^{32}$P]ATP and polynucleotide kinase and checked for binding of the relevant GST/FREAC protein in a gelshift assay (see below). Products from the last amplification step were gel-purified, digested with *Eco*RI and *Xba*I and cloned between the corresponding sites in pBluescript SK⁻. Approximately 30 inserts were sequenced for each FREAC protein.

### *DNase I footprinting*

pBluescript recombinants containing selected binding sites were linearized with *Xho*I, dephosphorylated with calf intestine alkaline phosphatase and then either 5′-labelled with [γ-$^{32}$P]ATP and polynucleotide kinase or 3′-labelled with [α-$^{32}$P]dATP and Klenow. The probes were cut out from the plasmid with *Sac*I and gel purified. 20 000 c.p.m. Cerenkov of probe were incubated with 10 or 100 ng FREAC – GST fusion protein in 50 μl 1× BB containing 1 μg poly(dIC) and 2% polyvinylalcohol. DNase I digestion and work-up procedure followed the method of Jones *et al.* (1988).

### *Expression of SWAP proteins*

To generate chimeras between FREAC-3 and FREAC-4, we combined the mutagenesis method of Nelson and Long (1989) with the E-PCR method of Kain *et al.* (1991) using the following primers: 5′-AGGGCCATCATCCCCTTCGCC (SWAP-1 and -5), 5′-ACACCGCC-TACGACGACAAGACGG (SWAP-2 and -6), 5′-AGGCCCAGGTCGC-AGGTCATC (SWAP-3 and -7), 5′-GGCGGCGTCCTTCGACGGCAA (SWAP-4 and -8), 5′-CCAAGCTTCTAATACGACTCACTA (T7-24mer), 5′-CTAAGGTCCTATGGGCGCCGCTGACGCGCCCTGACG (GST-36mer) and 5′-CCAAGCTTCTAATACGACTCACTATAGGGTT-TTTATTTTTAATTTTCTTTCAAATACTTCCACCATGGCCACGTTT-GGTGGTGGCGAC (T7-88mer). The first PCR step used T7-88mer and the relevant SWAP primer on a pGEX-KG/FREAC plasmid template (FREAC-3 for SWAP-5 to -8 and FREAC-4 for SWAP-1 to -4). Products were gel-purified and extended on the opposite pGEX-KG/FREAC plasmid template, after which the full-length chimeras with T7 promoter were amplified with GST-36mer and T7-24mer. After gel purification of the PCR products were transcribed *in vitro* with T7 RNA polymerase and the RNAs were translated in a reticulocyte lysate in the presence of [$^{35}$S]methionine. An aliquot from each translation was analysed on 15% SDS–PAGE and the yield was calculated by the excision of bands from the gel and quantitation by liquid scintillation counting. To express the parental FREAC proteins, the T7-88mer and GST-36mer primers were used directly on the respective pGEX-KG/FREAC plasmid templates.

### *Gelshift assay*

All gelshift gels, except for those used in the circular permutation assay, were 6% polyacrylamide (29:1) and 5% glycerol in Tris–glycine buffer (25 mM Tris, 190 mM glycine, 1 mM EDTA) and were run at 10–15 V/cm at +4°C for 90–120 min. Binding reactions (15 μl) contained 1× BB, 4 μg poly(dIC), typically 0.1 ng probe (20 000–50 000 c.p.m. Cerenkov) and protein as indicated in each case. Probes were generated by annealing complementary oligonucleotides and filling in 5′-GATC overhangs with Klenow enzyme and [α-$^{32}$P]dATP. Gelshift assays were quantitated by excision of bands and counting in a liquid scintillation counter or by scanning of autoradiograms followed by analysis with the ImageQuant software (Molecular Dynamics).

### *Circular permutation assay*

Double-stranded oligonucleotides A, B, E, F (sequences shown in Figure 4B) and G (5′-GATCCAAGTAAATAAACAATAGATC) were cloned into the *Bgl*II site of pCY4 (Prentki *et al.*, 1987). Probes were generated by digestion with *Eco*RI, *Hin*dIII, *Bst*NI, *Eco*RV, *Nhe*I or *Bam*HI, followed by dephosphorylation, 5′-labelling with [γ-$^{32}$P]ATP and polynucleotide kinase, and purification on PAGE. Gelshifts were performed according to Thompson and Landy (1988) on gels containing 5% glycerol, 0.5× TBE and 6, 8 or 10% polyacrylamide (29:1). Binding reactions were performed as described above, except that ~2 ng probe were used. Mobilities of protein–DNA complexes were measured and the ratio between the fastest (*Bam*HI) and the slowest (*Eco*RV) was used to calculate the bending angle according to Thompson and Landy (1988).

## Acknowledgements

We would like to thank Kerstin Dahlenborg and Majlis Hermansson for excellent technical assistance, and Gunnar Bjursell and Per Sunnerhagen for comments on this manuscript. This work was supported by grants from the Swedish Medical Research Council, the Knut and Alice Wallenberg Foundation and the Inga-Britt and Arne Lundberg Foundation. GenBank accession numbers for FREAC1–7 are U13219–13225, respectively.

## References

Altaba,A.R., Cox,C., Jessell,T.M. and Klar,A. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 8268–8272.

Bork,P., Ouzounis,C., Sander,C., Scharf,M., Schneider,R. and Sonnhammer,E. (1992) *Protein Sci.*, **1**, 1677–1690.

Brennan,R.G. (1993) *Cell*, **74**, 773–776.

Carlsson,P., Waterman,M.L. and Jones,K.A. (1993) *Genes Dev.*, **7**, 2418–2430.

Catron,K.M., Iler,N. and Abate,C. (1993) *Mol. Cell. Biol.*, **13**, 2354–2365.

Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294–5299.

Clark,K.L., Halay,E.D., Lai,E. and Burley,S.K. (1993) *Nature*, **364**, 412–420.

Clevidence,D.E., Overdier,D.G., Tao,W., Qian,X., Pani,L., Lai,E. and Costa,R.H. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 3948–3952.

Costa,R.H., Grayson,D.R. and Darnell,J.,Jr (1989) *Mol. Cell. Biol.*, **9**, 1415–1425.

Dirksen,M.L. and Jamrich,M. (1992) *Genes Dev.*, **6**, 599–608.

Echelard,Y., Epstein,D.J., St-Jacques,B., Shen,L., Mohler,J., McMahon,J.A. and McMahon,A.P. (1993) *Cell*, **75**, 1417–1430.

Ekker,S.C., Young,K.E., von Kessler,D.P. and Beachy,P.A. (1991) *EMBO J.*, **10**, 1179–1186.

Ekker,S.C., von Kessler,D.P. and Beachy,P.A. (1992) *EMBO J.*, **11**, 4059–4072.

Enerbäck,S., Ohlsson,B.G., Samuelsson,L. and Bjursell,G. (1992) *Mol. Cell. Biol.*, **12**, 4622–4633.

Genetics Computer Group (1991) *Program Manual for the GCG Package*. Genetics Computer Group, Madison, WI.

Giese,K. and Grosschedl,R. (1993) *EMBO J.*, **12**, 4667–4676.

Giese,K., Cox,J. and Grosschedl,R. (1992) *Cell*, **69**, 185–195.

Grosschedl,R., Giese,K. and Pagel,J. (1994) *Trends Genet.*, **10**, 94–100.

Grossniklaus,U., Pearson,R.K. and Gehring,W.J. (1992) *Genes Dev.*, **6**, 1030–1051.

Gruenberg,D.A., Natesan,S., Alexandre,C. and Gilman,M.Z. (1992) *Science*, **257**, 1089–1095.
Guan,K.L. and Dixon,J.E. (1991) *Anal. Biochem.*, **192**, 262–267.
Häcker,U., Grossniklaus,U., Gehring,W.J. and Jäckle,H. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 8754–8758.
Herschlag,D. and Johnson,F.B. (1993) *Genes Dev.*, **7**, 173–179.
Jones,K.A., Luciw,P.A. and Duchange,N. (1988) *Genes Dev.*, **2**, 1101–1114.
Kaestner,K.H., Lee,K.H., Schlondorff,J., Hiemisch,H., Monaghan,A.P. and Schütz,G. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 7628–7631.
Kaestner,K.H., Hiemisch,H., Luckow,B. and Schütz,G. (1994) *Genomics*, **20**, 377–385.
Kain,K.C., Orlandi,P.A. and Lanar,D.E. (1991) *Biotechniques*, **10**, 366–374.
Lai,E., Prezioso,V.R., Tao,W.F., Chen,W.S. and Darnell,J.Jr (1991) *Genes Dev.*, **5**, 416–427.
Lai,E., Clark,K.L., Burley,S.K. and Darnell,J.Jr (1993) *Proc. Natl Acad. Sci. USA*, **90**, 10421–10423.
Lamb,P. and McKnight,S.L. (1991) *Trends Biochem. Sci.*, **16**, 417–422.
Li,C., Lai,C.F., Sigman,D.S. and Gaynor,R.B. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 7739–7743.
Li,C., Lusis,A.J., Sparkes,R., Nirula,A. and Gaynor,R. (1992a) *Genomics*, **13**, 665–671.
Li,C., Lusis,A.J., Sparkes,R., Tran,S.M. and Gaynor,R. (1992b) *Genomics*, **13**, 658–664.
Lin,Y.S., Carey,M., Ptashne,M. and Green,M.R. (1990) *Nature*, **345**, 359–361.
Miller,L.M., Gallegos,M.E., Morisseau,B.A. and Kim,S.K. (1993) *Genes Dev.*, **7**, 933–947.
Moitoso de Vargas,L., Kim,S. and Landy,A. (1989) *Science*, **244**, 1457–1461.
Natesan,S. and Gilman,M.Z. (1993) *Genes Dev.*, **7**, 2497–2509.
Nelson,R.M. and Long,G.L. (1989) *Anal. Biochem.*, **180**, 147–151.
Overdier,D.G., Porcella,A. and Costa,R.H. (1994) *Mol. Cell. Biol.*, **14**, 2755–2766.
Prentki,P., Pham,M.H. and Galas,D.J. (1987) *Nucleic Acids Res.*, **15**, 10060.
Riddle,R.D., Johnson,R.L., Laufer,E. and Tabin,C. (1993) *Cell*, **75**, 1401–1416.
Sasaki,H. and Hogan,B.L. (1993) *Development*, **118**, 47–59.
Sasaki,H. and Hogan,B.L. (1994) *Cell*, **76**, 103–115.
Schleif,R. (1992) In McKnight,S.L. and Yamamoto,K.R. (eds), *Transcriptional Regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 643–665.
Struhle,U., Blader,P., Henrique,D. and Ingham,P.W. (1993) *Genes Dev.*, **7**, 1436–1446.
Thompson,J.F. and Landy,A. (1988) *Nucleic Acids Res.*, **16**, 9687–9705.
Travis,A., Amsterdam,A., Belanger,C. and Grosschedl,R. (1991) *Genes Dev.*, **5**, 880–894.
Tsuchiya,S., Yamabe,M., Yamaguchi,Y., Kobayashi,Y., Konno,T. and Tada,K. (1980) *Int. J. Cancer*, **26**, 171–176.
Waterman,M.L., Fischer,W.H. and Jones,K.A. (1991) *Genes Dev.*, **5**, 656–669.
Weigel,D. and Jäckle,H. (1990) *Cell*, **63**, 455–456.
Weigel,D., Jurgens,G., Kuttner,F., Seifert,E. and Jäckle,H. (1989) *Cell*, **57**, 645–658.
Wu,H.M. and Crothers,D.M. (1984) *Nature*, **308**, 509–513.
Zhu,G., Muller,E.G., Amacher,S.L., Northrop,J.L. and Davis,T.N. (1993) *Mol. Cell. Biol.*, **13**, 1779–1787.
Zinkel,S.S. and Crothers,D.M. (1991) *J. Mol. Biol.*, **219**, 201–215.

*Received on June 13, 1994; revised on July 27, 1994*